Sensitive data guidelines

Authors with sensitive data, or other data that cannot be shared openly, should apply appropriate restrictions before sharing their data. Authors should put their data in a repository (advantageous for transparency, long-term storage and managing access requests) where possible and only fully restrict data access if no other sharing option is available.

Sensitive or restricted data includes:

- **Identifiable human data:** data involving human research participants may present a risk of reidentification if shared openly. This includes both quantitative and qualitative research data. A list of identifiers in human data is available here.
- **Other sensitive data:** non-human data sensitivities should also be considered, from locations of endangered species or archaeologically sensitive areas, to sensitive military or governmental data.
- **Proprietary/third party data:** authors must ensure they have the necessary rights and permissions before sharing data via any method. Copyright and data ownership should be considered separately to sensitivities related to research participants. Authors should identify with any restrictions as early as possible and establish what, if any, sharing is permitted. See this guidance from the UK Data Service.

If any of the above are applicable, authors should consider the following methods to facilitate safe sharing of sensitive data. Participant consent to share the data (in addition to use or collection of data) should also be obtained and documented prior to data collection.

- Anonymising data to create a shareable version. It may be possible to remove or replace identifying information in the data before sharing openly.
  See guidance on anonymisation from *Trials* and the UK Data Service.

- Use of a controlled access repository to manage who can access data and under what terms. Certain repositories offer this functionality, for example enabling a data owner to know who has access to the data and/or to apply additional restrictions such as an agreement not to reidentify participants. re3data.org provides functionality to search for controlled-access repositories and content types. Additionally, this blog provides advice on searching for controlled-access repositories and this article provides guidance on suitable repositories. While controlled-access agreements tend to require each potential user to be specifically approved, some repositories, such as OpenICPSR, instead employ a blanket end-user agreement to protect participants from identification. Other services such as ClinicalStudyDataRequest.com can facilitate interactions between the data generator(s) and those applying for access.

- Use of Trusted Research Environments or data safe havens. Certain research institutions manage environments within which data can be queried and accessed by trusted parties only, without removing data from the system. These are mainly associated with clinical health settings. Contact your research institute to check if this is an option for your data.

- Use of metadata records in repositories for data that cannot be publicly shared. This provides persistent, long-term context to another researcher on what data are available, even if the data can only be made available on request. See this example from *npj Precision Oncology*, and this blog for more details.